ELSEVIER

Stochastic Processes and their Applications 52 (1994) 1–16

stochastic processes and their applications

Nonparametric estimators for Markov step processes[†]

P.E. Greenwood[a], W. Wefelmeyer[b,*]

[a] *Department of Mathematics, University of British Columbia, Vancouver, B.C., Canada V6T 1Y4*
[b] *Department of Mathematics, University of Siegen, 57068 Siegen, Germany*

Received 5 October 1992; Revised 27 August 1993

Abstract

The distribution of a homogeneous, continuous-time Markov step process with values in an arbitrary state space is determined by the transition distribution and the mean holding time, which may depend on the state. We suppose that both are unknown, introduce a class of functionals which determines the transition distribution and the mean holding time up to equivalence, and construct estimators for the functionals. Assuming that the embedded Markov chain is Harris recurrent and uniformly ergodic, and that the mean holding time is bounded and bounded away from 0, we show that the estimators are asymptotically efficient, as the observation time increases. Then we consider the two submodels in which the mean holding time is assumed constant, and constant and known, respectively. We describe efficient estimators for the submodels. For finite state space, our results give efficiency of an estimator for the generator which was studied by Lange (1955) and Albert (1962).

Key words: Efficient estimator; Markov step process; Nonparametric estimation

1. Introduction

Statistical inference for *parametric* Markov step process models has recently attracted considerable attention. Here we treat *nonparametric* models. For a certain class of functionals we introduce simple estimators and prove their efficiency. Let us first recall related results for i.i.d. observations and for (discrete-time) Markov chains.

Suppose we observe i.i.d. realizations $X_1, \ldots, X_n$ from an unknown distribution $P(\mathrm{d}x)$ on an arbitrary state space. The distribution P is determined by expectations Pf assigned to a sufficiently rich class of bounded functions $f(x)$. The *empirical estimator* for Pf is $n^{-1}\sum_{j=1}^{n} f(X_j)$; it is efficient in nonparametric models (Levit, 1974, 1975).

*Corresponding author.

[†]Work supported by NSERC, Canada.

SSDI 0304-4149(93)E0068-P

Now suppose that our observations $X_0, \dots, X_n$ come from a Harris recurrent and uniformly ergodic Markov chain with unknown transition distribution $Q(x, \mathrm{d}y)$ and invariant distribution $\pi(\mathrm{d}x)$. The transition distribution Q is determined π-almost surely by expectations

$$\pi Q f = \int\int \pi(\mathrm{d}x) Q(x, \mathrm{d}y) f(x, y)$$

assigned to certain bounded functions $f(x, y)$ under the invariant joint distribution of two successive observations. The functional $\pi Q f$ corresponds exactly to Pf in the i.i.d. case. It admits a simple estimator, the *empirical estimator* $n^{-1} \sum_{j=1}^{n} f(X_{j-1}, X_j)$. The estimator is efficient (Greenwood and Wefelmeyer, 1992, extending Penev, 1991).

Are there analogous results for continuous-time processes? The simplest such process to describe is a Markov step process $X = (X_t)_{t \geqslant 0}$. It stays in state x for an exponentially distributed time with mean $\lambda(x)^{-1}$, then jumps to state y according to a transition distribution $Q(x, \mathrm{d}y)$. As for Markov chains the transition distribution Q is determined π-almost surely by expectations $\pi Q f$. Here $\pi(\mathrm{d}x) Q(x, \mathrm{d}y)$ is the invariant joint distribution of two successive observations from the embedded Markov chain $(X_{T_j})_{j \geqslant 0}$, with T_j denoting the successive jump times of X. The mean holding time function λ^{-1} is determined π-almost surely by expectations

$$\pi(\lambda^{-1} f) = \int \pi(\mathrm{d}x) \lambda(x)^{-1} f(x)$$

for certain bounded functions $f(x)$. The functional is natural because the measure $\pi(\mathrm{d}x)\lambda(x)^{-1}$ is proportional to the invariant distribution of the process X. Again, $\pi Q f$ and $\pi(\lambda^{-1} f)$ admit simple, efficient estimators. That is our main result. It can be described more explicitly as follows.

Suppose we observe a Markov step process X on the time interval $[0, n]$. The choice $[0, n]$ is convenient for asymptotic statements. Everything remains true for intervals $[0, T]$. Let $j(n)$ denote the observed number of jumps. Set $T_0 = 0$. Assume that the state space is separable. Let Q and λ be unknown, with Q Harris recurrent and uniformly ergodic, and λ bounded and bounded away from 0. Then, as the observation time n tends to infinity,

$$j(n)^{-1} \sum_{j=1}^{j(n)} f(X_{T_{j-1}}, X_{T_j}) \text{ is efficient for } \pi Q f,$$

$$j(n)^{-1} \int_0^n f(X_t)\,\mathrm{d}t \text{ is efficient for } \pi(\lambda^{-1} f).$$

Efficiency of the first estimator follows from a martingale approximation similar to the one for Markov chains in Greenwood and Wefelmeyer (1992). To prove efficiency of the second estimator, we express it as compensator of a process for which we can, again, obtain a martingale approximation. Here efficiency is meant in the sense of a nonparametric version of the Hájek–LeCam convolution theorem for regular estimators, Proposition 4.1.

The reader who finds the results heuristically obvious may want to test his intuition with similar estimators involving functions f with different arguments, like

$$(j(n)-1)^{-1}\sum_{j=2}^{j(n)} f(X_{T_{j-2}}, X_{T_{j-1}}, X_{T_j}) \quad \text{and} \quad j(n)^{-1}\sum_{j=1}^{j(n)} (T_j - T_{j-1}) f(X_{T_{j-1}}, X_{T_j}).$$

These estimators are consistent for the functionals

$$\int\int\int \pi(\mathrm{d}x) Q(x, \mathrm{d}y) Q(y, \mathrm{d}z) f(x, y, z) \quad \text{and} \quad \int\int \pi(\mathrm{d}x)\lambda(x)^{-1} Q(x, \mathrm{d}y) f(x, y),$$

respectively. They are, however, not efficient. We omit the calculations. A heuristic argument for inefficiency of the first estimator in the case of Markov chains is in Greenwood and Wefelmeyer (1992).

The paper is organized as follows. The setting is described in Section 2. An infinite-dimensional version of local asymptotic normality for Markov step processes is described in Section 3, an asymptotic variance bound for regular estimators of functionals of λ and Q in Section 4. Section 5 contains the main results. We consider in particular the two submodels in which λ is assumed constant, and constant and known, respectively. We also consider the case when the state space is finite. Then the model has a finite-dimensional parameter, the generator. An estimator suggested by Lange (1955) was shown to attain the Cramér–Rao bound by Albert (1962). We complement the result by showing that the estimator is also efficient in a stronger sense. For a direct proof see Höpfner (1988). The proofs are collected in Section 6.

2. The setting

Consider an arbitrary state space E with countably generated σ-field $\mathscr{E}$. Let X denote the canonical process on the space of all right continuous, piecewise constant functions mapping $[0, \infty)$ into E, with finitely many jumps in finite time. Introduce the *jump times* $T_0 = 0$ and

$$T_j = \inf\{t > T_{j-1}: X_t \neq X_{T_{j-1}}\},$$

and the *canonical filtration* $\mathscr{F}_t = \sigma(X_s: s \leq t)$, $t \geq 0$. Let $\mathscr{F}$ be the σ-field generated by $\mathscr{F}_t$, $t \geq 0$.

Fix a transition distribution $Q(x, \mathrm{d}y)$ which is Harris recurrent and uniformly ergodic, and a function $\lambda(x)$ on the state space which is positive, bounded and bounded away from 0, and write

$$G(x, \mathrm{d}y) = \lambda(x) Q(x, \mathrm{d}y).$$

Then for each x in E there exists a unique probability measure P_x on $\mathscr{F}$ such that X is a *Markov step process* starting from x with G as *generator*. This means that $X_0 = x$

almost surely, $(X_{T_j})_{j \geq 0}$ is a Markov chain with transition distribution Q, and conditionally on $\mathscr{F}_{T_j}$ the *holding time* $T_{j+1} - T_j$ of the process X in state X_{T_j} is exponential with mean $\lambda(X_{T_j})^{-1}$.

Let $\pi(\mathrm{d}x)$ denote the invariant distribution of Q. Recall that a measure $m(\mathrm{d}x)$ is called *invariant* for X if, for all A in $\mathscr{E}$ and $t \geq 0$,

$$\int m(\mathrm{d}x) P_x(X_t \in A) = m(A).$$

We have the following relation between π and m:

$$m(\mathrm{d}x) = (\pi\lambda^{-1})^{-1}\pi(\mathrm{d}x)\lambda(x)^{-1}. \tag{2.1}$$

The relation implies

$$m(\mathrm{d}x)G(x, \mathrm{d}y) = (\pi\lambda^{-1})^{-1}\pi(\mathrm{d}x)Q(x, \mathrm{d}y). \tag{2.2}$$

For functions $f'(x)$ of one argument we use the standard notation

$$mf' = \int m(\mathrm{d}x)f'(x),$$

$$G_x f' = \int G(x, \mathrm{d}y)f'(y),$$

$$mGf' = \int\int m(\mathrm{d}x)G(x, \mathrm{d}y)f'(y).$$

It will be convenient to use similar notation for functions $f(x, y)$ of two arguments,

$$G_x f = \int G(x, \mathrm{d}y)f(x, y).$$

3. Local asymptotic normality

In this section we describe a convenient infinite-dimensional version of local asymptotic normality for Markov step process models. We start the process with a *fixed* initial distribution. The results remain true as long as the initial distributions are not assumed to depend strongly on the generator. In particular, they remain true if the processes are stationary. Let $\mathscr{G}$ denote the family of all generators on the state space E. Fix a generator G for which X is positive Harris recurrent and uniformly ergodic, and for which λ is bounded and bounded away from 0. Consider the function space

$$H = \{h : E^2 \to \mathbb{R} \text{ measurable, bounded}\}.$$

It will play the role of a *local parameter space* at G. We can write each function $h(x, y)$ in H as a sum of two functions, one depending only on x, the other having conditional expectation 0 given x under Q,

$$h'(x) = Q_x h, \qquad h''(x, y) = h(x, y) - Q_x h.$$

Correspondingly, $H = H' + H''$. For h' and h'' we introduce *local parametrizations* of λ and Q, respectively, around G,

$$\lambda_{n,h'}(x) = \lambda(x)(1 + n^{-1/2} h'(x)),$$

$$Q_{nh''}(x, \mathrm{d}y) = Q(x, \mathrm{d}y)(1 + n^{-1/2} h''(x, y)).$$

We write

$$G_{nh}(x, \mathrm{d}y) = \lambda_{nh'}(x) Q_{nh''}(x, \mathrm{d}y).$$

If G governs the process, we write P for the distribution of the process; if G_{nh} governs the process, we write P^{nh}. As usual, P_t denotes the restriction of P to $\mathscr{F}_t$.

It will be convenient to associate with X a multivariate point process on $[0, \infty) \times E \times E$,

$$\mu(\mathrm{d}t, \mathrm{d}x, \mathrm{d}y) = \sum_{j \geq 1} \varepsilon_{(T_j, X_{T_{j-1}}, X_{T_j})}(\mathrm{d}t, \mathrm{d}x, \mathrm{d}y).$$

Here ε_a stands for the Dirac measure with mass in point a. The conditional distribution of $(T_j, X_{T_{j-1}}, X_{T_j})$ given $\mathscr{F}_{T_{j-1}}$ is

$$1_{(T_{j-1}, T_j]}(t)\, \varepsilon_{X_{T_{j-1}}}(\mathrm{d}x) \lambda(x) \exp(-(t - T_{j-1})\lambda(x))\, \mathrm{d}t\, Q(x, \mathrm{d}y).$$

We obtain an explicit representation of the compensator of μ from Jacod (1975) or Jacod and Shiryaev (1987, p. 136, Theorem 1.33) as

$$\nu(\mathrm{d}t, \mathrm{d}x, \mathrm{d}y) = \sum_{j \geq 1} 1_{(T_{j-1}, T_j]}(t)\, \mathrm{d}t\, \varepsilon_{X_{T_{j-1}}}(\mathrm{d}x) G(x, \mathrm{d}y).$$

For the number of jumps observed up to time n write

$$j(n) = \max\{j \colon T_j \leq n\}.$$

For f in H introduce the integrals

$$f * \mu_n = \int_0^n \int_{E \times E} f(x, y) \mu(\mathrm{d}t, \mathrm{d}x, \mathrm{d}y) = \sum_{j=1}^{j(n)} f(X_{T_{j-1}}, X_{T_j}), \tag{3.1}$$

$$f * \nu_n = \int_0^n \int_{E \times E} f(x, y) \nu(\mathrm{d}t, \mathrm{d}x, \mathrm{d}y) = \int_0^n \int_E G(X_t, \mathrm{d}y) f(X_t, y)\, \mathrm{d}t. \tag{3.2}$$

With the notations introduced above, we can state *local asymptotic normality* in concise form.

Proposition 3.1. *For h in H,*

$$\log \mathrm{d}P_n^{nh}/\mathrm{d}P_n = n^{-1/2}\, h*(\mu-\nu)_n - \tfrac{1}{2}\, mGh^2 + o_P(1), \tag{3.3}$$

$$n^{-1/2}\, h*(\mu-\nu)_n \Rightarrow N_h \quad \text{under } P. \tag{3.4}$$

Here N_h is a normal random variable with mean 0 and variance mGh^2.

A proof of Proposition 3.1 can be obtained from *parametric* results for Markov step processes. For discrete state space, Höpfner (1988) has conditions for local asymptotic normality. For arbitrary state space, Höpfner (1993a, 1993b) gives conditions for likelihoods to be locally asymptotically mixed normal and locally asymptotically quadratic, and obtains local asymptotic normality as a special case.

We note that weaker conditions suffice for weak convergence to a Gaussian shift model rather than a *stochastic* approximation as in (3.3); see Höpfner et al. (1990). However, we prove efficiency of estimators via a *stochastic* characterization in Proposition 4.2 below which requires local asymptotic normality, not just weak convergence to a Gaussian shift model.

The norm $(mGh^2)^{1/2}$ in local asymptotic normality determines how difficult it is, asymptotically, to distinguish between P_n and P_n^{nh}. The norm provides an inner product $mG(h\bar{h})$ on the local parameter space. Note that H' and H'' are *orthogonal* with respect to the inner product. In particular, using relations (2.1) and (2.2) between the invariant distributions of the process and the embedded chain,

$$\begin{aligned} mG(h\bar{h}) &= m(\lambda h'\bar{h}') + mG(h''\bar{h}'') \\ &= (\pi\lambda^{-1})^{-1}(\pi(h'\bar{h}') + \pi Q(h''\bar{h}'')). \end{aligned} \tag{3.5}$$

Choice of local parametrization: Here we have considered λ and Q as parameters of the model. We could also take the generator as parameter since λ and Q are determined by G,

$$\lambda(x) = G(x, E), \qquad Q(x, \mathrm{d}y) = G(x, y)/G(x, E).$$

Then the local parameters would be the functions h in H, and the local parametrization could for instance be introduced as

$$G_{nh}^*(x, \mathrm{d}y) = G(x, \mathrm{d}y)(1 + n^{-1/2} h(x, y)).$$

This would lead to a slightly more concise description of the local model. However, in Section 5 it will be convenient to work with h' and h'' rather than h. For example, we will look at submodels given by certain restrictions on λ. For the local parametrization $\lambda_{nh'}$ these restrictions translate immediately into restrictions on the local parameter h'.

It is true that we can obtain local parametrizations of λ and Q from G_{nh}^*, via

$$\lambda_{nh}(x) = G_{nh}^*(x, E) = \lambda(x)(1 + n^{-1/2} Q_x h),$$

$$Q_{nh}(x, \mathrm{d}y) = G_{nh}^*(x, \mathrm{d}y)/G_{nh}^*(x, E) = 1 + n^{-1/2}(h(x, y) - Q_x h) + O(n^{-1}).$$

But then Q_{nh} depends on h, not just on h'', and an additional error $O(n^{-1})$ appears. These problems are harmless, but unnecessary, and we avoid them.

4. An asymptotic variance bound

In this section we recall briefly an asymptotic variance bound for estimators of functionals of the generator in the setting of Section 3. We choose a submodel among all Markov step processes considered here. The *full* model is described by the family $\mathscr{G}$ of all generators on E. We consider an arbitrary submodel. It is described by a subfamily $\mathscr{G}_0$. We fix $\mathscr{G}$ in $\mathscr{G}_0$ with the properties listed in Section 3, and recall that generators G_{nh} were defined there. We choose a subspace H_0 in H, the *local parameter space of* $\mathscr{G}_0$ at G, such that $G_{nh} \in \mathscr{G}_0$ for $h \in H_0$. The *local model* at G at time n is then given by the generators G_{nh}, $h \in H_0$.

Let k be a real-valued functional on $\mathscr{G}_0$. Call k *differentiable* at G in $\mathscr{G}_0$ with *gradient* g in H_0 if

$$n^{1/2}(k(G_{nh}) - k(G)) \to mG(hg) \quad \text{for } h \text{ in } H_0.$$

Let $\hat{k}_n$ be a sequence of real-valued estimators. Call $\hat{k}_n$ regular for k at G in $\mathscr{G}_0$ with *limit* L if

$$n^{1/2}(\hat{k}_n - k(G_{nh})) \Rightarrow L \quad \text{under } P^{nh} \text{ for } h \text{ in } H_0.$$

With these definitions, we have an asymptotic variance bound.

Proposition 4.1. *If k is differentiable at G in $\mathscr{G}_0$ with gradient g in H_0, and $\hat{k}_n$ is regular for k at G in $\mathscr{G}_0$ with limit L, then*

$$L = N_g + M \quad \text{in distribution,}$$

with M independent of N_g.

As in Section 3 we denote by N_g a normal random variable with mean 0 and variance mGg^2. Proposition 4.1 justifies calling $\hat{k}_n$ *efficient* for k at G in $\mathscr{G}_0$ if $L = N_g$ in distribution. By (3.5) we can write the variance as

$$mGg^2 = (\pi\lambda^{-1})^{-1}(\pi g'^2 + \pi Q g''^2).$$

Choice of local model: We have not tried hard to make the local model as large as possible, and there may be no estimator that attains the lower bound on the risk *globally*, i.e. for each possible underlying G in $\mathscr{G}_0$. However:

(1) Trying hard is no guarantee. Drastic examples of globally unattainable bounds are in Ritov and Bickel (1990).

(2) The local model enters the definition of a regular estimator, and unnecessarily large local models, containing for example *all* smooth paths through G, may rule out competing estimators.

(3) The specific local models in Section 5 lead to bounds which we show to be globally attainable.

We recall a useful characterization of efficient and regular estimators. Call an estimator $\hat{k}_n$ *asymptotically linear* for k at G with *influence function* f in H if

$$n^{1/2}(\hat{k}_n - k(G)) = n^{-1/2} f * (\mu - \nu)_n + o_P(1).$$

This is analogous to the well-known definition in the i.i.d. case in that one uses an expression of the form appearing in local asymptotic normality, (3.3), for approximating the standardized error of the estimator.

Proposition 4.2. *An estimator $\hat{k}_n$ is regular and efficient for k at G in $\mathcal{G}_0$ if and only if it is asymptotically linear for k at G with influence function the gradient of k at G in $\mathcal{G}_0$.*

In Section 5 we use Proposition 4.2 to prove efficiency of estimators. A convenient reference for Propositions 4.1 and 4.2 is Greenwood and Wefelmeyer (1990); asymptotic linearity is described through relation (2.18) there.

5. The results

This section contains our main results. The proofs are in Section 6. We consider first the full model described by the family $\mathcal{G}$ of all generators on E, and fix a generator G as in Section 3. We will describe efficient estimators for the functionals πQf and $\pi(\lambda^{-1} f')$. We begin with πQf. First we calculate its gradient.

Lemma 5.1. *For f in H, the functional πQf is differentiable at G in $\mathcal{G}$ with gradient $(\pi\lambda^{-1})Af$ in H. Here*

$$(Af)(x, y) = f(x, y) - Q_x f + \sum_{k=0}^{\infty} (Q_y^k Qf - Q_x^{k+1} Qf)$$

is a linear operator which maps H into H'' and is the identity on H''.

A version of Lemma 5.1 for Markov chains and functions $f(x, y) = f'(y)$ of one argument was first obtained by Penev (1991). To prove asymptotic linearity of the estimator described in Theorem 5.4 below, we use the following two lemmas. The first, simple, lemma will be used several times.

Lemma 5.2. *The sequence $n^{-1} j(n)$ converges to $m\lambda = (\pi\lambda^{-1})^{-1}$ in P-probability.*

The next lemma implies a martingale approximation for an empirical estimator, (5.1) below. A version for Markov chains is in Greenwood and Wefelmeyer (1992).

Lemma 5.3. *Uniformly for uniformly bounded f in H,*

$$(f - \pi Qf - Af) * \mu_n = \mathrm{O}_P(\log n).$$

Because Af is in H'', we have $Q_x Af = 0$ for all x in E. Hence Lemmas 5.2 and 5.3 imply that the estimator $j(n)^{-1} f * \mu_n$ is asymptotically linear for πQf at G, with influence function the gradient in H:

$$n^{1/2}(j(n)^{-1} f * \mu_n - \pi Qf) = n^{-1/2}(\pi\lambda^{-1})(Af) * (\mu - \nu)_n + \mathrm{o}_P(1). \tag{5.1}$$

Our first main result now follows immediately from the characterization of efficient estimators given in Proposition 4.2.

Theorem 5.4. *Fix f in H. The estimator*

$$j(n)^{-1} f * \mu_n = j(n)^{-1} \sum_{j=1}^{j(n)} f(X_{T_{j-1}}, X_{T_j})$$

is regular and efficient for πQf at G in $\mathscr{G}$. Its asymptotic variance is

$$(\pi\lambda^{-1})\pi Q(Af)^2.$$

Suppose now that more is assumed known about λ, say that it is constant or that it is fixed, or both. Then the local parameter space $H = H' + H''$ reduces to a smaller space $H_0 = H_0' + H''$. Since the gradient of πQf is in H'', it remains in H_0. Hence $j(n)^{-1} f * \mu_n$ remains efficient.

Functional version: Suppose we want to estimate a function of the form $f \to \pi Qf$, with f running through some index class of bounded functions, and we want to obtain efficiency in a functional sense, i.e. for the function-valued estimator $f \to j(n)^{-1} f * \mu_n$ regarded as a process indexed by f. Then we need tightness of the sequence of *empirical processes*

$$f \to n^{1/2} j(n)^{-1}(f - \pi Qf) * \mu_n.$$

Since $(X_{T_{j-1}}, X_{T_j})$ is φ-mixing with exponential rate, tightness, and hence functional efficiency, follows from a functional central limit theorem for mixing sequences. A recent reference is Arcones and Yu (1994).

We turn to the functional $\pi(\lambda^{-1} f')$. First we calculate its gradient.

Lemma 5.5. *For f' in H' the functional $\pi(\lambda^{-1} f')$ is differentiable at G in $\mathscr{G}$ with gradient in H given by*

$$(\pi\lambda^{-1})(-\lambda(x)^{-1} f'(x) + A(\lambda^{-1} f')(x, y)).$$

The operator A is defined in Lemma 5.1; *here it is applied to*

$$f(x, y) = \lambda(y)^{-1} f'(y).$$

Our second main result uses Lemmas 5.2, 5.3 and 5.5. The proof is in Section 6.

Theorem 5.6. *Fix f' in H'. The estimator $j(n)^{-1}\int_0^n f'(X_t)\,\mathrm{d}t$ is regular and efficient for $\pi(\lambda^{-1}f')$ at G in $\mathscr{G}$. Its asymptotic variance is*

$$(\pi\lambda^{-1})(\pi(\lambda^{-1}f')^2+\pi Q(A(\lambda^{-1}f'))^2).$$

The basic idea of the proof is the following. To prove asymptotic linearity of the estimator, one interprets $\int_0^\bullet f'(X_t)\,\mathrm{d}t$ as compensator of the process $(\lambda^{-1}f')*\mu$ and uses the martingale approximation of Lemma 5.3 for this process.

The distribution of the process is determined by functionals of the form πQf and $\pi(\lambda^{-1}f')$ because they admit simple estimators. Another family of functionals with this property is given by mGf and mf'. Efficient estimators for these functionals are easily obtained, once noticed (see (2.1) and (2.2)) that

$$mGf=(\pi\lambda^{-1})^{-1}\pi Qf,$$

$$mf'=(\pi\lambda^{-1})^{-1}\pi(\lambda^{-1}f').$$

By Theorem 5.6, applied for $f'=1$,

$n^{-1}j(n)$ is regular and efficient for $(\pi\lambda^{-1})^{-1}$ at G in $\mathscr{G}$.

Hence Theorems 5.4 and 5.6 imply that

$$n^{-1}f*\mu_n \text{ is regular and efficient for } mGf \text{ at } G \text{ in } \mathscr{G}, \tag{5.2}$$

$$n^{-1}\int_0^n f'(X_t)\,\mathrm{d}t \text{ is regular and efficient for } mf' \text{ at } G \text{ in } \mathscr{G}. \tag{5.3}$$

Finite state space: For finite state space, the generator reduces to a matrix. The matrix may be viewed as a finite-dimensional parameter of the model. Albert (1962) has shown that the maximum likelihood estimator attains the Cramér–Rao bound. We express the maximum likelihood estimator in terms of our 'nonparametric' estimators and conclude that it is also efficient in a stronger sense. A direct proof is due to Höpfner (1988).

Let $E=\{1,\ldots,m\}$. The generator G is determined by the matrix with entries $G_{pq}=G(p,\{q\})$. The invariant measure is described by the vector with components $m_p=m(\{p\})$. Fix p, q in E with $p\neq q$. We want to estimate G_{pq}. For

$$f(x,y)=1_{(p,q)}(x,y) \quad \text{and} \quad f'(x)=1_p(x),$$

the functionals mGf and mf' reduce to

$$mGf=m_pG_{pq} \quad \text{and} \quad mf''=m_p.$$

A regular and efficient estimator of m_pG_{pq} is obtained from (5.2) as $n^{-1}N_{pq}(n)$, with $N_{pq}(n)$ the number of transitions from p to q observed up to time n.

$$N_{pq}(n)=\sum_{j=1}^{j(n)}1_{(p,q)}(X_{T_{j-1}},X_{T_j})=f*\mu_n.$$

A regular and efficient estimator for m_p is obtained from (5.3) as $n^{-1}S_p(n)$, with $S_p(n)$ the time spent in p up to time n,

$$S_p(n) = \int_0^n 1_p(X_t)\,\mathrm{d}t = \int_0^n f'(X_t)\,\mathrm{d}t.$$

This implies that

$S_p(n)^{-1}N_{pq}(n)$ is regular and efficient for G_{pq} at G in $\mathscr{G}$.

Lange (1955) and Albert (1962) obtain this estimator as a maximum likelihood estimator.

Constant mean holding time: Consider the submodel described by the family $\mathscr{G}_0$ of all generators $G(x, \mathrm{d}y) = \lambda Q(x, \mathrm{d}y)$ with *constant* mean holding time $1/\lambda$. Then the local parameter space reduces to $H_0 = \mathbb{R} + H''$. As noted after Theorem 5.4, the estimator $j(n)^{-1} f * \mu_n$ remains efficient for πQf in the submodel. Here it suffices to consider the functional λ rather than $\pi(\lambda^{-1}f')$.

Theorem 5.7. *The estimator $n^{-1}j(n)$ is regular and efficient for λ at G in $\mathscr{G}_0$. Its asymptotic variance is λ.*

The estimator for $\pi(\lambda^{-1}f')$ in Theorem 5.6 has an influence function which is not in H_0. Hence it is *not* efficient in $\mathscr{G}_0$, except when m and hence π is a one-point measure or f' is constant. To obtain an efficient estimator for $\pi(\lambda^{-1}f')$ in $\mathscr{G}_0$, we write

$$\pi(\lambda^{-1}f') = \lambda^{-1}\pi f'$$

and find efficient estimators for λ and $\pi f'$. Their ratio is then efficient for $\pi(\lambda^{-1}f')$. By Theorem 5.7, the estimator $n^{-1}j(n)$ is efficient for λ in $\mathscr{G}_0$. An efficient estimator for $\pi f'$ in the full model $\mathscr{G}$ is obtained from Theorem 5.4, applied for $f(x, y) = f'(y)$:

$j(n)^{-1} \sum_{j=1}^{j(n)} f'(X_{T_j})$ is regular and efficient for $\pi f'$ at G in $\mathscr{G}$.

By the remark following Theorem 5.4, the estimator remains efficient in the submodel $\mathscr{G}_0$. Hence

$nj(n)^{-2} \sum_{j=1}^{j(n)} f'(X_{T_j})$ is regular and efficient for $\pi(\lambda^{-1}f')$ at G in $\mathscr{G}_0$.

Constant and known mean holding time: Consider the submodel described by the family $\mathscr{G}_1$ of all generators $G(x, \mathrm{d}y) = Q(x, \mathrm{d}y)$, with mean holding time $1/\lambda(x)$ equal to 1 for all x in E. Then it suffices to consider functionals πQf. For them, $j(n)^{-1} f * \mu_n$ is efficient in $\mathscr{G}_1$ by Theorem 5.4 and the remark following it.

Since $n^{-1}j(n)$ converges to 1 in probability by Lemma 5.2, we have

$$n^{1/2}(j(n)^{-1}\; f * \mu_n - \pi Qf) = n^{-1/2} \sum_{j=1}^{n} (f(X_{T_{j-1}}, X_{T_j}) - \pi Qf) + o_P(1).$$

In other words, estimating πQf from the path of the Markov step process on $[0, n]$ is asymptotically equivalent to estimating πQf from $n + 1$ observations of a Markov chain. Hence efficiency of $j(n)^{-1} f * \mu_n$ in $\mathscr{G}_1$ also follows from the efficiency result for Markov chains in Greenwood and Wefelmeyer (1992).

6. Proofs

Lemma 5.1 gives the gradient of the functional πQf. To calculate the gradient, we need to know how π depends on Q. To this end, we recall some results on Markov processes. Since Q is uniformly ergodic, there is an a in $(0, 1)$ such that

$$\| Q^k - \Pi \| \leq a^k \quad \text{for } k \geq 1, \tag{6.1}$$

with $\Pi(x, \mathrm{d}y) = \pi(\mathrm{d}y)$ the *stationary projection* of Q. Here the norm of a transition kernel $K(x, \mathrm{d}y)$ is the operator norm

$$\| K \| = \sup\{ \| \mu K \| : \| \mu \| \leq 1 \},$$

with $\| \mu \|$ the variation norm of a signed measure $\mu(\mathrm{d}x)$.

For transition distributions $\bar{Q}$ on E with invariant distributions $\bar{\pi}$ we have (Kartashov, 1985) the von Neumann expansion

$$\bar{\pi} - \pi = \pi(\bar{Q} - Q)R + \mathrm{o}(\| \bar{Q} - Q \|) \tag{6.2}$$

with

$$R = I + \sum_{k=1}^{\infty} (Q^k - \Pi). \tag{6.3}$$

Here $I(x, \mathrm{d}y) = \varepsilon_x(\mathrm{d}y)$ is the *identity kernel.*

Proof of Lemma 5.1. Since $\Pi_y Qf = \pi Qf$ does not depend on y, the sum

$$\sum_{k=0}^{\infty} (Q_y^k Qf - Q_x^{k+1} Qf)$$

converges by inequality (6.1). By definition of $Q_{nh''}$,

$$\| Q_{nh''} - Q \| = \mathrm{O}(n^{-1/2}).$$

Hence relation (6.2) implies

$$\| \pi_{nh''} - \pi \| = \mathrm{O}(n^{-1/2}). \tag{6.4}$$

Fix h'' in H'' and set $K(x, \mathrm{d}y) = Q(x, \mathrm{d}y)\, h''(x, y)$. By relations (6.2)–(6.4),

$$n^{1/2}(\pi_{nh''} Q_{nh''} f - \pi Qf) \to \pi Kf + \pi KQf + \pi K \sum_{k=1}^{\infty} (Q^k - \Pi)\, Qf.$$

Again by inequality (6.1), the infinite sum on the right side can be approximated by a finite one. Since $K_x 1 = Q_x h'' = 0$ for all x,

$$\pi K(Q^k - \Pi)Qf = \pi K Q^k Qf$$
$$= \int\int \pi(\mathrm{d}x) K(x, \mathrm{d}y)(Q_y^k Qf - Q_x^{k+1} Qf).$$

By definition of K and Af we obtain

$$n^{1/2}(\pi_{nh''} Q_{nh''} f - \pi Q f) \to \pi Q(h'' Af). \tag{6.5}$$

Expressing this in terms of the inner product (3.5), we see that the gradient is $(\pi\lambda^{-1})Af$.

Proof of Lemma 5.2. Write

$$n^{-1} j(n) = n^{-1} 1 * \mu_n = n^{-1} 1 * \nu_n + n^{-1} 1 * (\mu - \nu)_n.$$

By the martingale central limit theorem (3.4),

$$n^{-1} 1 * (\mu - \nu)_n = \mathrm{o}_P(n^{-1/2}).$$

By relation (3.2) and the ergodic theorem,

$$n^{-1} 1 * \nu_n = n^{-1} \int_0^n \lambda(X_t)\,\mathrm{d}t = m\lambda + \mathrm{o}_P(1).$$

The result follows.

Proof of Lemma 5.3. Let $\varepsilon > 0$. By Lemma 5.2 there exists $b > 0$ such that for all n,

$$P\{|n^{-1} j(n) - m\lambda| > b\} < \varepsilon.$$

The following relations hold uniformly for $|f| \leq 1$ and paths for which

$$|n^{-1} j(n) - m\lambda| \leq b.$$

By inequality (6.1), for c sufficiently large,

$$\left| \sum_{k > c \log n} \sum_{j=1}^{j(n)} (Q_{X_{T_j}}^k Qf - Q_{X_{T_{j-1}}}^{k+1} Qf) \right| \leq 2n(m\lambda + b) \sum_{k > c \log n} a^k \to 0.$$

Hence

$$\sum_{j=1}^{j(n)} (Af)(X_{T_{j-1}}, X_{T_j}) = \sum_{j=1}^{j(n)} (f(X_{T_{j-1}}, X_{T_j}) - Q_{X_{T_{j-1}}} f)$$
$$+ \sum_{0 \leq k \leq c \log n} \sum_{j=1}^{j(n)} (Q_{X_{T_j}}^k Qf - Q_{X_{T_{j-1}}}^{k+1} Qf) + \mathrm{o}(1).$$

Rearranging the sums, we see that the conditional expectations cancel except for $j = 0$ and $j = j(n)$. Hence the right side equals

$$\sum_{j=1}^{j(n)} (f(X_{T_{j-1}}, X_{T_j}) - Q_{X_{T_{j-1}}}^{(c \log n)+1} Qf) + \sum_{0 \le k \le c \log n} Q_{X_{T_{j(n)}}}^k Qf - Q_{X_0}^k Qf).$$

The second sum is of order $\log n$ since f is bounded. The result now follows by replacing $Q_{X_{T_{j-1}}}^{(c \log n)+1} Qf$ by πQf. The error is negligible by a second application of inequality (6.1),

$$\left| \sum_{j=1}^{j(n)} (Q_{X_{T_{j-1}}}^{(c \log n)+1} Qf - \pi Qf) \right| \le n(m\lambda + b) a^{(c \log n)+1} \to 0.$$

Proof of Lemma 5.5. Since λ is bounded away from 0, we have uniformly for x in E,

$$\lambda_{nh'}(x) - \lambda(x) = n^{-1/2} \lambda(x) h'(x) = O(n^{-1/2})$$

and

$$\begin{aligned} \lambda_{nh'}(x)^{-1} - \lambda(x)^{-1} &= -\lambda(x)^{-2}(\lambda_{nh'}(x) - \lambda(x)) + O(n^{-1}) \\ &= -n^{-1/2} \lambda(x)^{-1} h'(x) + O(n^{-1}). \end{aligned}$$

Together with (6.4),

$$\pi_{nh''}(\lambda_{nh'}^{-1} f') - \pi(\lambda^{-1} f') = -n^{-1/2} \pi(\lambda^{-1} h' f') + (\pi_{nh''} - \pi)(\lambda^{-1} f') + O(n^{-1}).$$

From (6.5), applied for $f(x, y) = \lambda(y)^{-1} f'(y)$,

$$n^{1/2}(\pi_{nh''}(\lambda_{nh'}^{-1} f') - \pi(\lambda^{-1} f')) \to -\pi(h' \lambda^{-1} f') + \pi Q(h'' A(\lambda^{-1} f')).$$

The gradient is now obtained by expressing the right side in terms of the inner product (3.5).

Proof of Theorem 5.6. Compare (3.1) and (3.2) and note that

$$\int_0^{\bullet} f'(X_t)\, dt = (\lambda^{-1} f') * \nu \tag{6.6}$$

is the compensator of $(\lambda^{-1} f') * \mu$. The proof now reduces to an application of Lemmas 5.3 and 5.2. Write

$$(\lambda^{-1} f') * \nu_n = -(\lambda^{-1} f') * (\mu - \nu)_n + (\lambda^{-1} f') * \mu_n. \tag{6.7}$$

By Lemma 5.3, applied for $f(x, y) = \lambda(y)^{-1} f'(y)$,

$$(\lambda^{-1} f' - \pi(\lambda^{-1} f') - A(\lambda^{-1} f')) * \mu_n = O_P(\log n). \tag{6.8}$$

Recall that the operator A maps into the space H'' consisting of bounded functions $h''(x, y)$ with $Q_x h'' = 0$ for x in E. Hence we can write

$$A(\lambda^{-1} f') * \mu_n = A(\lambda^{-1} f') * (\mu - \nu)_n. \tag{6.9}$$

Relations (6.6)–(6.9) and Lemma 5.2 imply

$$n^{1/2}(j(n))^{-1}\int_0^n f'(X_t)\,dt - \pi(\lambda^{-1}f'))$$

$$= n^{-1/2}(\pi\lambda^{-1})(-\lambda^{-1}f' + A(\lambda^{-1}f'))*(\mu - \nu)_n + o_P(1).$$

This means that the estimator is asymptotically linear for $\pi(\lambda^{-1}f')$ at G. By Lemma 5.5, the influence function is the gradient of $\pi(\lambda^{-1}f')$ at G in H. Theorem 5.6 now follows from the characterization of efficient estimators given in Proposition 4.2.

Proof of Theorem 5.7. First we calculate the gradient for λ at G in $\mathscr{G}_0$. By (3.5) the inner product on $H_0 = \mathbb{R} + H''$ is

$$mG(h\bar{h}) = \lambda(h'\bar{h}' + \pi Q(h''\bar{h}'')).$$

The functional λ is differentiable,

$$n^{1/2}(\lambda_{nh'} - \lambda) = \lambda h'.$$

Expressing the right-hand side in terms of the inner product $mG(h\bar{h})$, we see that the gradient of λ in H_0 equals 1. By relations (3.1) and (3.2),

$$n^{1/2}(n^{-1}j(n) - \lambda) = n^{-1/2}\,1*(\mu - \nu)_n.$$

Hence the estimator $n^{-1}j(n)$ is asymptotically linear at G with influence function equal to 1, the gradient of λ in H_0. Theorem 5.7 now follows from the characterization of efficient estimators in Proposition 4.2.

Acknowledgement

We are most grateful to the referees for eliminating some inaccuracies and suggesting improvements in the presentation.

References

A. Albert, Estimating the infinitesimal generator of a continuous time, finite state Markov chain, Ann. Math. Statist. 33 (1962) 727–753.

M.A. Arcones and B. Yu, Central limit theorems for empirical processes of stationary mixing sequences, J. Theoret. Probab. 7 (1994) 47–71.

P.E. Greenwood and W. Wefelmeyer, Efficiency of estimators for partially specified filtered models, Stochastic Process Appl. 36 (1990) 353–370.

P.E. Greenwood and W. Wefelmeyer, Efficiency of empirical estimators for Markov chains, Preprints in Statistics 135, Mathematical Institute, University of Cologne (1992).

R. Höpfner, Asymptotic inference for continuous-time Markov chains, Probab. Theory Rel. Fields 77 (1988) 537–550.

R. Höpfner, Asymptotic inference for Markov step processes: observation up to a random time, Stochastic Process. Appl. 48 (1993a) 295–310.

R. Höpfner, On statistics of Markov step processes: representation of log-likelihood ratio processes in filtered local models, Probab. Theory Rel. Fields 94 (1993b) 375–398.

R. Höpfner, J. Jacod and L. Ladelli, Local asymptotic normality and mixed normality for Markov statistical models, Probab. Theory Rel. Fields 86 (1990) 105–129.

J. Jacod, Multivariate point processes: predictable projection, Radon-Nikodym derivatives, representation of martingales, Z. Wahrsch. verw. Gebiete 31 (1975) 235–253.

J. Jacod and A.N. Shiryaev, Limit Theorems for Stochastic Processes (Springer, New York, 1987).

N.V. Kartashov, Criteria for uniform ergodicity and strong stability of Markov chains with a common state space, Theory Probab. Math. Statist. 30 (1985) 71–89.

O. Lange, Statistical investigation of parameters in Markov processes, Colloq. Math. 3 (1955) 147–160.

B.Ya. Levit, On optimality of some statistical estimates, in: J. Hájek, ed., Proceedings of the Prague Symposium on Asymptotic Statistics, Vol. 2 (Charles University, Prague, 1974).

B.Ya. Levit, On the efficiency of a class of non-parametric estimates, Theory Probab. Appl. 20 (1975) 723–740.

S. Penev, Efficient estimation of the stationary distribution for exponentially ergodic Markov chains, J. Statist. Plann. Inference 27 (1991) 105–123.

Y. Ritov and P.J. Bickel, Achieving information bounds in non and semiparametric models, Ann. Statist. 18 (1990) 925–938.